

# Linkage between sexual orientation and chromosome Xq28 in males but not in females

Stella Hu<sup>1</sup>, Angela M. L. Pattatucci<sup>1</sup>, Chavis Patterson<sup>1</sup>, Lin Li<sup>1</sup>, David W. Fulker<sup>2</sup>, Stacey S. Cherny<sup>2</sup>, Leonid Kruglyak<sup>3</sup> & Dean H. Hamer<sup>1</sup>

We have extended our analysis of the role of the long arm of the X chromosome (Xq28) in sexual orientation by DNA linkage analyses of two newly ascertained series of families that contained either two gay brothers or two lesbian sisters as well as heterosexual siblings. Linkage between the Xq28 markers and sexual orientation was detected for the gay male families but not for the lesbian families or for families that failed to meet defined inclusion criteria for the study of sex-linked sexual orientation. Our results corroborate the previously reported linkage between Xq28 and male homosexuality in selected kinships and suggest that this region contains a locus that influences individual variations in sexual orientation in men but not in women.

<sup>1</sup>Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

<sup>2</sup>Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309, USA

<sup>3</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02139, USA

Correspondence should be addressed to D.H.H.

Most men are sexually attracted to females whereas most women are sexually attracted to males. The mechanisms underlying the development of this pronounced dimorphism in heterosexual orientation are largely unknown, but probably involve a complex interaction between genetic, biological, experiential and socio-cultural factors. Equally unknown are the origins of intragender variations in sexual orientation such as homosexuality and bisexuality<sup>1-4</sup>.

Recent experiments have suggested that one variation in sexual orientation, namely predominant or exclusive homosexuality in males, is influenced by a genetic locus on the long arm of the X chromosome<sup>5</sup>. The initial suggestion of sex-linkage came from family pedigree studies which showed that homosexual men had more gay male relatives through maternal than through paternal lineages; control experiments showed that these gay male probands reported approximately equal numbers of female homosexual relatives on both sides of the family, and that lesbian probands reported approximately equal numbers of male homosexual relatives on both sides of the family, indicating that the observed maternal loading was not an artifact of differential knowledge or reporting of maternal compared to paternal gay relatives<sup>4,6</sup>. The main evidence for sex-linkage came from DNA linkage analysis of a selected series of families in which there were two known gay brothers and a lack of evidence for nonmaternal or non-sex-limited transmission of homosexuality. A significant excess of these selected gay male sib-pairs were concordant for DNA markers on chromosome Xq28, suggesting that this region contains genetic information that increases the probability of homosexual orientation<sup>5</sup>.

The current work had three objectives. First, to test the hypothesis of a linkage between Xq28 and male homosexuality in a new set of selected gay male sib-pair families. This was especially important in view of the poor reliability of some of the previously reported DNA linkages to psychological traits<sup>7</sup>. Second, to study the segregation of the Xq28 region in the heterosexual brothers of homosexual male sib-pairs. This was done to estimate the penetrance of the hypothesized Xq28 locus and to test for segregation distortion in this region. Third, to determine whether Xq28 is correlated with sexual orientation in women. There have been relatively few studies on the role of inheritance in female sexual orientation, and the question of whether male and female homosexuality are aetiologically overlapping or distinct is unresolved<sup>6,8-13</sup>. Our new results support the presence of an Xq28 locus influencing male, but not female, sexual predisposition.

## Experimental design and families

Our protocol was designed to study male sexual orientation in families with two gay brothers and female sexual orientation in families with two lesbian sisters. We chose this strategy, rather than analyzing families with both male and female homosexuals, because it would detect a sex-common locus even if male and female homosexuality involved sex-specific alleles or different environmental factors.

The basic criterion for entry in the study was the presence of two same-sex, non-twin siblings with a predominant or exclusive homosexual orientation as judged by sexual history and the Kinsey scales<sup>4-5,14-15</sup>. For the male study, families in which there was direct

Table 1 Allele sharing in sib-pairs

Locus	Homosexual sibs		Heterosexual sibs		Combined $t_{tot}^e$
	$\bar{\pi}_{gg} \pm SE^a$	$t_{gg}^b$	$\bar{\pi}_{gs} \pm SE^c$	$t_{gs}^d$	
<b>a, Males<sup>f</sup></b>					
<b>Xq27</b>					
A....DXS548	.499 ± .079	ns			
B....FRAXA	.553 ± .086	ns			
A/B	.567 ± .085	ns	.273 ± .141	ns	ns
<b>Xq28</b>					
C....DXS52	.638 ± .077	1.78*			
D....G6PD	.543 ± .055	ns			
E....F8C	.624 ± .070	1.77*			
F....DXS1108	.670 ± .077	2.22*			
G....JXYQ28	.574 ± .077	ns			
C/D/E/F/G	.671 ± .081	2.09*	.219 ± .115	2.44*	3.00**
<b>Xq/Yq</b>					
H....DXYS154	.534 ± .087	ns	.260 ± .134	ns	ns
<b>b, Females<sup>g</sup></b>					
<b>Xq27</b>					
A....DXS548	.364 ± .122	ns			
B....FRAXA	.655 ± .099	ns			
A/B	.614 ± .100	ns	.500 ± .189	ns	ns
<b>Xq28</b>					
C....DXS52	.501 ± .100	ns			
F....DXS1108	.573 ± .094	ns			
C/F	.584 ± .093	ns	.556 ± .176	ns	ns
<b>Xq/Yq</b>					
H....DXYS154	.492 ± .095	ns	.625 ± .183	ns	ns

\*  $P < 0.05$  (one-sided t-test); \*\*  $P < 0.005$  (one-sided t-test); ns  $P > 0.05$  (one-sided t-test); <sup>a</sup>  $\bar{\pi}_{gg}$  is the estimated proportion of genes shared by-descent ( $\pm$  standard error of the mean) for homosexual-homosexual sib-pairs; <sup>b</sup>  $t_{gg}$  is the t-score for a test of  $\bar{\pi}_{gg} > 0.5$ ; <sup>c</sup>  $\bar{\pi}_{gs}$  is the estimated proportion of genes shared by-descent ( $\pm$  standard error of the mean) for a heterosexual sibling with his or her marker-concordant homosexual siblings; <sup>d</sup>  $t_{gs}$  is the t-score for a test of  $\bar{\pi}_{gs} < 0.5$ ; <sup>e</sup>  $t_{tot}$  is the t-score for a test of  $\bar{\pi}_{gg} > \bar{\pi}_{gs}$ ; <sup>f</sup> Male study:  $n = 33$  homosexual-homosexual sib-pairs, 12 homosexual-heterosexual sib-pairs; <sup>g</sup> Female study:  $n = 36$  homosexual-homosexual sib-pairs, 11 homosexual-heterosexual sib-pairs.

evidence for nonmaternal transmission of homosexuality (that is, families in which the father, son or paternal half-brother of the proband was nonheterosexual) were excluded because they would not be consistent with the hypothesis of X-chromosome linkage. For the female study this sort of selection was not feasible since women receive X chromosomes from both parents. Individuals who identified as bisexual, or who gave ambiguous or contradictory sexual histories or interview responses, were not included in the primary analysis because of diagnostic uncertainty. Families in which there were three or more nonheterosexual same-sex siblings were also treated separately to conform to the initial analysis and because of the complications in ascertaining and analyzing such kinships<sup>5</sup>. (The numbers and reasons for the exclusions in the male and female studies are described in Methods.)

The male study included 33 gay male sib-pair families that were unrelated to one another and to the previously described study population<sup>5</sup> and met the entrance criteria. Eight of these families included heterosexual brothers from whom DNA samples and interview or questionnaire results were available. In addition, four heterosexual brothers of the previously

reported gay male sib-pair families were used to analyse the penetrance of the locus. The female study involved 36 lesbian sib-pair families that met the entrance criteria, 15 of which included heterosexual sisters from whom DNA and personal information was obtained.

DNA samples from family members, which included mothers and other relatives whenever possible, were typed for a series of markers spanning the distal long arm of the X chromosome from the DXS548 locus in Xq27 through the DXYS154 locus in the Xq/Yq homology region at the telomere. The inheritance of the Xq28 DNA markers was determined through identity-by-descent (IBD) in 78% of the potentially informative families and through identity-by-state (IBS) in the remaining cases.

### Male study: included families

**Gay sib-pairs.** Our first aim was to test for linkage between Xq28 and homosexual orientation in a new series of gay male sib-pairs. The test hypothesis was that more than 50% of these predominantly or exclusively homosexual brothers would share DNA markers in this region. Of the 33 eligible sib-pairs, 32 were potentially informative for the five closely spaced markers in Xq28 that were analysed. Twenty-two out of these 32 sib-pairs shared all of their Xq28 markers. This gave an estimated level of haplotype sharing, taking into account the IBS cases, of 67% (Table 1a), which was significantly greater than the null hypothesis of 50% ( $P < 0.05$  by one-sided t-test;  $P = 0.0384$  by simulation test on  $n = 10,000$  replicates). Multipoint maximum likelihood analysis<sup>16</sup> gave a similar estimate of 68% gene sharing in the Xq28 region. Lower and non-significant levels of allele sharing were observed for the markers in Xq27 and the Xq/Yq homology region.

**Heterosexual brothers.** Our second goal was to analyse the segregation of the Xq28 region in heterosexual siblings. The test hypothesis was that less than 50% of the heterosexual brothers of Xq28-concordant gay sib-pairs would share markers with their homosexual siblings. The brothers of gay sib-pairs discordant for Xq28 marker alleles were not included in this analysis because they are intrinsically non-informative with regard to penetrance or segregation distortion. There were 11 families included in the analysis: seven new families from the selected cohort described above (Fig. 1, families M1–M7) and four previously analysed families for which DNA was available from a previously unreported heterosexual brother.

Most of the heterosexual brothers had different Xq28 markers from their homosexual siblings. In nine families, the single heterosexual brother was genetically discordant with his gay brothers; in one family the heterosexual brother was genetically concordant-by-descent with his homosexual siblings; and in one family, for which maternal DNA was not available, the two heterosexual brothers were genetically concordant-by-state with their concordant-by-state homosexual siblings. These data gave an estimated level of Xq28 haplotype sharing of 22% (Table 1a), which was less than the null hypothesis of 50% ( $P < 0.05$  by one-sided t-test;  $P \approx 0.02$ – $0.07$  by exact binomial test for 10/12 or 9/12 discordant pairs). Multipoint analysis<sup>16</sup> gave a similar estimate of 24% gene sharing in the Xq28

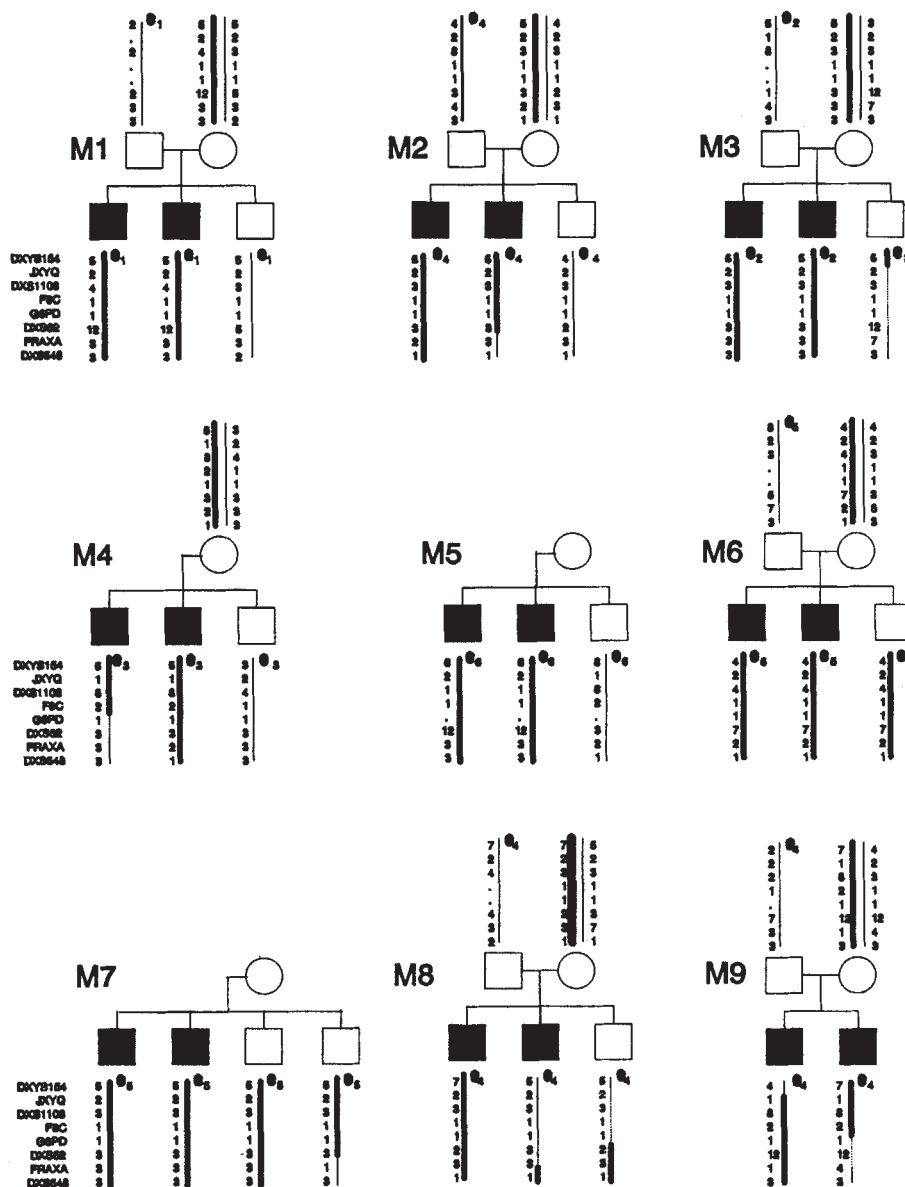


Fig. 1 Selected male families. Shown are the DNA marker genotyping results for all new gay sib-pair families for which DNA from a heterosexual brother was available (M1–M8) and a family with two potentially informative recombination events (M9). ■ homosexual male; □ heterosexual male; ○ heterosexual female. Only mothers, genotyped sons and genotyped fathers are indicated. Haplotypes are those that minimize the number of recombinants. Changes in the thickness of the haplotype bars indicate recombination events and dashed lines indicate ambiguity in the breakpoint due to maternal marker homozygosity. The short bars represent the Xq/Yq homology region of the Y chromosome. Families M1–M7, in which the gay siblings share Xq28 marker alleles, were included in the analysis of penetrance (Table 1a); family M8, in which the gay brothers are discordant for the Xq28 markers, was not.

region for these sib pairs. The data for the Xq27 region and for the Xq/Yq marker were not significant.

The overall evidence for a linkage between male sexual orientation and Xq28 was evaluated by examining the difference between the proportion of alleles shared for the homosexual–homosexual and homosexual–heterosexual sib-pairs. The homosexual–homosexual pairs were more likely to have the same Xq28 marker alleles than were the homosexual–heterosexual pairs (Table 1a;  $P < 0.005$  by one-sided t-tests of haplotype data and of multipoint regression data;  $P = 0.0175$  by a conservative estimate using Fisher's method for combining independent  $P$  values). Again, the data for

markers outside Xq28 were statistically insignificant.

**Multipoint interval mapping.** In an attempt to refine the localization of the putative X-linked male sexual orientation related locus, we use a modified sib-pair linkage analysis<sup>17</sup> which analyses all available markers simultaneously using a regression-based interval mapping procedure<sup>18–20</sup> adapted for X linkage. Furthermore, to explore the relationship between the Kinsey subscales and the putative quantitative trait locus (QTL), we carried out the analysis on the individual scales of sexual self-identification, attraction, fantasy, and activity, as well as on the average Kinsey score and the categorical measure of sexual orientation. All possible pairs of sibs within a family were used, and the sample size was corrected for non-independence among the pairs within a single family<sup>21</sup>. Only the newly ascertained families within the selected cohort were included in the analysis.

The regression-based interval mapping results are shown in Fig. 2 as a plot of  $t$ -values versus genetic map position. All four Kinsey subscales, their average, and the categorical measure of sexual orientation yielded qualitatively similar conclusions, with the maximum linkage occurring coincident with the *DXS52* locus ( $t_{\max} = 2.93$  for average Kinsey scale rating;  $P < 0.005$  by one-sided t-test;  $P < 0.02$  after conservative Bonferroni correction for independent markers<sup>20</sup>). Analysis of the data by an alternative multipoint maximum likelihood procedure<sup>16</sup> gave precisely the same peak position and similar significance levels ( $t_{\max} = 2.95$  for average Kinsey scale rating; see Methods). Interestingly, the quantitative Kinsey scales generated slightly higher test statistics than the qualitative data. Simulation

experiments using the multipoint maximum likelihood approach gave significance levels of  $P = 0.009$  ( $n = 1,005$ ) for average Kinsey score,  $P = 0.0124$  ( $n = 1,777$ ) for the sexual attraction scale, and  $P = 0.037$  ( $n = 4,763$ ) for sexual orientation as a dichotomous trait. Although these results indicate that the most likely position of the sexual orientation QTL is close to *DXS52*, the sample size is too small to exclude statistically the remainder of the mapped region.

**X-specific and X/Y homologous sequences**

The telomeric 320 kilobases (kb) of the long arms of the X and Y chromosomes are homologous to one



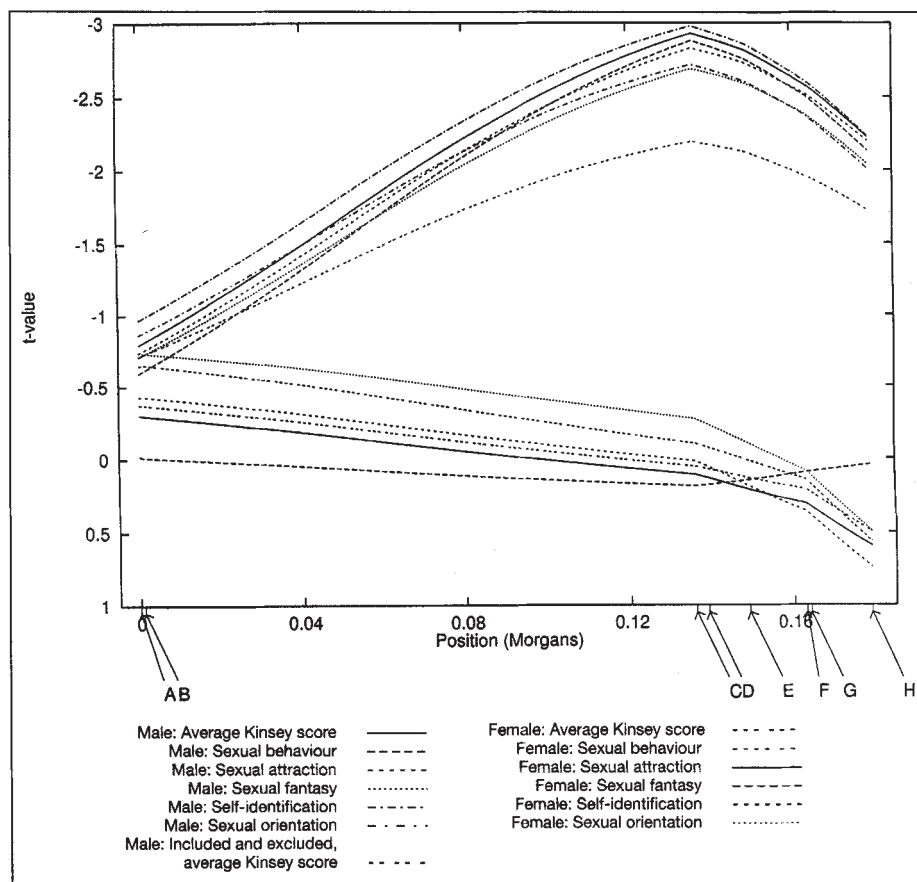


Fig. 2 Multipoint interval mapping of the Kinsey scales of sexual orientation in males and females. The letters indicate the positions of the marker loci listed in Table 1. This analysis includes all possible sib-pairs from the newly ascertained gay male families that were included in the main study ('Male',  $n = 52$  pairs), these families plus the gay male families that were excluded from the main study ('Male: Included and Excluded, Average Kinsey Score',  $n = 74$  pairs), and the 36 lesbian families that were included in the main study ('Female',  $n = 78$  pairs).

another and can undergo sequence exchanges in both male and female meioses<sup>22,23</sup>. Inspection of the DNA marker pedigrees revealed just two potentially informative maternal recombination events between the X-specific and Xq/Yq-homologous markers in our current set of gay male sib-pair families (Fig. 1). In family M9, the two homosexual siblings were genetically discordant at *DXYS154* but were concordant-by-descent at the Xq28 loci including *JXYQ28*, a newly available marker that precisely demarcates the boundary between the X-specific and Xq/Xq homologous sequences. In family M3, a heterosexual brother inherited the same *DXYS154* marker as his homosexual siblings but had a different allele at the X-specific locus *DXS52*. Both of these families are consistent with the multipoint mapping of the sexual orientation related locus to the X-specific portion of Xq28.

No unequivocal recombinants between markers within the X-specific portion of Xq28 were observed. There were multiple recombinants between the Xq28 and Xq27 regions, which are separated by 13.5 cM in our map; most of these were consistent with the sexual orientation locus lying in Xq28 (such as family M2), but there were also some families consistent with an Xq27 localization, presumably due to heterogeneity (such as family M8). In families in which the precise position of a recombination event was unclear due to maternal marker homozygosity, the most likely posi-

tion of the crossover was estimated from the flanking marker data<sup>19</sup>; for example, the two homosexual sons in families M4 and M9 have a > 90% chance of sharing *DXS52* (which was non-informative in these families) because they are identical-by-descent for *F8C* (located 1.3 cM towards the telomere) even though they are discordant for *FRAXA* (located 13.5 cM towards the centromere).

#### Male study: excluded families

There were seven families with gay or bisexual/ambiguous brothers that failed to meet the entrance criteria and were therefore excluded from the main study prior to initiating genotype analysis. These excluded families were genotyped as part of other ongoing studies, and the data is shown for purposes of comparison. The extent of Xq28 haplotype sharing for these families was assessed according to two models: A) that individuals who identified as bisexual or who gave ambiguous or inconsistent sexual histories or interview responses were in fact homosexual; and B) that such individuals were in fact heterosexual. Table 2 shows that the levels of Xq28 haplotype sharing were not significantly greater or less than 50% for either the homosexual-homosexual or homosexual-het-

erosexual pairs as defined by either model, and therefore this small dataset could neither confirm nor refute linkage under either of the test hypotheses. When the included and excluded groups were combined to give a total sample, statistical significance in favour of linkage was retained ( $P < 0.05$ ), but the level of homosexual-homosexual gene sharing was decreased and the level of homosexual-heterosexual gene sharing was increased in comparison to the selected cohort.

The data for the excluded families was also analysed by QTL multipoint mapping, which treats sexual orientation as a continuous rather than dimorphic trait. The excluded pairs gave no evidence for linkage to any region of the map (not shown). Fig. 2 shows that when the excluded families were added to the main sample, the shape and peak position for the multipoint curve remained unaltered. However, the statistical evidence for linkage decreased from  $t_{\max} = 2.93$  ( $P \approx 0.003$ ) to  $t_{\max} = 2.20$  ( $P \approx 0.016$ ).

#### Female study

To determine whether female sexual orientation is correlated with Xq28, we performed similar DNA linkage studies on a series of 36 lesbian sib-pair families. There was no significant evidence for excess sharing of any of the markers tested in the lesbian sib-pairs (Table 1b). Furthermore, the heterosexual sisters of Xq28-concor-

**Table 2 Xq28 haplotype sharing in included and excluded male sib-pairs**

Group <sup>a</sup>	Model <sup>b</sup>	Homosexual sibs			Heterosexual sibs			Total	
		N <sub>gg</sub> <sup>c</sup>	$\bar{\pi}_{gg} \pm SE^d$	t <sub>gg</sub> <sup>e</sup>	N <sub>gs</sub> <sup>f</sup>	$\bar{\pi}_{gs} \pm SE^g$	t <sub>gs</sub> <sup>h</sup>	N <sub>tot</sub> <sup>i</sup>	t <sub>tot</sub> <sup>j</sup>
Included	A/B	32	.671 ± .081	2.09*	12	.219 ± .115	2.44*	44	3.00**
Excluded	A	10	.491 ± .164	ns	4	.500 ± .289	ns	14	ns
Excluded	B	7	.566 ± .200	ns	6	.333 ± .211	ns	13	ns
Total	A	42	.628 ± .073	1.75*	16	.289 ± .112	1.89*	58	2.47*
Total	B	39	.652 ± .075	2.03*	18	.257 ± .101	2.40*	57	3.03**

\*  $P < 0.05$  (one-sided t-test); \*\*  $P < 0.005$  (one-sided t-test); ns  $P > 0.05$  (one-sided t-test); <sup>a</sup>Families were included or excluded from the male sib-pair study prior to genotype analysis; <sup>b</sup>In model A, individuals who identified as bisexual or who gave ambiguous or inconsistent sexual histories or interview responses were counted as homosexual. In model B, such individuals were counted as heterosexual; <sup>c</sup>N<sub>gg</sub> is the number of homosexual-homosexual sib-pairs; <sup>d</sup> $\bar{\pi}_{gg}$  is the estimated proportion of Xq28 haplotypes shared by-descent ( $\pm$  standard error of the mean) for homosexual-homosexual sib-pairs; <sup>e</sup>t<sub>gg</sub> is the t-score for a test of  $\bar{\pi}_{gg} > 0.5$ ; <sup>f</sup>N<sub>gs</sub> is the number of heterosexual-homosexual sib-pairs for families in which the homosexual siblings were Xq28-concordant; <sup>g</sup> $\bar{\pi}_{gs}$  is the estimated proportion of Xq28 haplotypes shared by-descent ( $\pm$  standard error of the mean) for a heterosexual sibling with his Xq28-concordant homosexual siblings; <sup>h</sup>t<sub>gs</sub> is the t-score for a test of  $\bar{\pi}_{gs} < 0.5$ ; <sup>i</sup>N<sub>tot</sub> is the total number of sib-pairs. <sup>j</sup>t<sub>tot</sub> is the t-score for a test of  $\bar{\pi}_{gg} > \bar{\pi}_{gs}$ .

dant sib-pairs had an apparently random possibility of inheriting the same or different markers as their lesbian sisters. Thus the combined data for the homosexual and heterosexual sisters was not significant for any of the X-chromosome markers or haplotypes examined.

The female subjects displayed more variability and less congruity in their Kinsey self-ratings than did the male subjects. Therefore the female data were further analysed by multipoint interval mapping of the Kinsey scales for sexual self-identification, attraction, fantasy, and behaviour, as well as average Kinsey score and the categorical trait of sexual orientation. Fig. 2 shows that none of these variables was significantly correlated to any region of the X chromosome that was examined.

The above analyses focused on the 36 lesbian sib-pair families that met the entrance criteria for the study. The 17 families that were excluded from the study also showed no significant linkage to Xq28. Combining the included and excluded families did not change this result.

**Discussion**

Sexual orientation is a complex characteristic that is influenced by many different factors. The multifactorial or heterogeneous nature of sexual orientation is clear from the observation that although the identical twin of a gay man or lesbian has an above average probability of being homosexual, he or she still has a 50% or more chance of identifying as heterosexual<sup>4</sup>. Therefore it is highly unlikely that any single genetic variation or allele will be present in all homosexual individuals or absent from all heterosexual individuals. Rather, our research focuses on the identification of loci that play a quantitative role in individual variations in sexual orientation.

The aim of this study was to test whether one particular locus, located at Xq28, plays such a quantitative role, and if so whether its effect is the same or different in men and women. Our study examines male-male and female-female homosexual sibling pairs and their heterosexual siblings. If Xq28 had no effect on sexual orientation, both the homosexual-homosexual and homosexual-heterosexual sib-pairs would be expected

to share on average 50% of their maternally inherited Xq28 DNA markers. By contrast, if Xq28 was involved in sexual orientation, the sharing of maternally inherited marker alleles would be expected to be greater than 50% for homosexual-homosexual sib-pairs and less than 50% for homosexual-heterosexual sib-pairs. By analysing separate groups of male-male and female-female sib-pairs, it was possible to compare the role of Xq28 in both male and female sexual orientation. The use of homosexual-heterosexual pairs provided a control for segregation distortion at the Xq28 region and an independent test for linkage.

Our data, summarized in Fig. 3, show a pattern of gene sharing that is consistent with a genetic influence of Xq28 in males but not in females. For males, the estimated probability of gene sharing for the X-specific sequences between JXYQ and DXS52 was 0.67, which was greater than the null hypothesis of 0.5 at a nominal significance level of  $P \approx 0.02-0.04$ . Because this experiment represents a single test of a previously stated hypothesis on a new set of families, it is not necessary to make any correction for multiple comparisons or non-independence and therefore represents a significant finding at the usual criterion of  $\alpha < 0.05$ . Furthermore, our results are not significantly different than those reported earlier for a different set of selected gay male sib-pairs<sup>5</sup>. These results suggest, in agree-

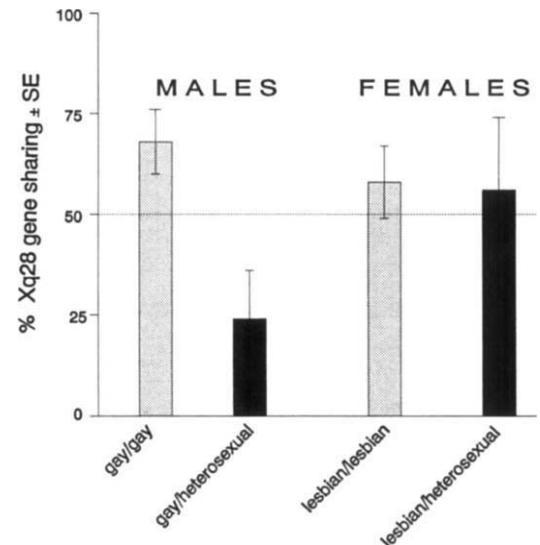


Fig. 3 Summary of gene sharing in male compared to female sib-pairs. The average levels of Xq28 gene sharing ( $\bar{\pi} \pm$  standard error of the mean) are shown for sib-pairs in which both members are homosexual and for sib-pairs in which one member is heterosexual and one member is homosexual and has the same Xq28 markers as his or her homosexual sibling. For males,  $n = 33$  homosexual-homosexual sib-pairs, 12 homosexual-heterosexual sib-pairs; for females,  $n = 36$  homosexual-homosexual sib-pairs, 11 homosexual-heterosexual sib-pairs.

ment with theoretical calculations and practical experience, that the shared-trait sib-pair method is an appropriate approach for the initial analysis of complex traits<sup>24–26</sup>.

Our previous research focused exclusively on homosexual brothers because the shared-trait sib-pair method is statistically and epidemiologically robust for detecting linkages to complex traits<sup>24</sup>. This design does not, however, yield any information about the penetrance of the locus, nor does it formally distinguish between linkage and segregation distortion (meiotic drive). Accordingly, we extended our analysis to include the heterosexual brothers of gay male sib-pairs. The estimated probability of a heterosexual brother sharing Xq28 markers with his marker-concordant gay brothers was 0.22, which was less than the null hypothesis of 0.5 ( $P \approx 0.02–0.07$ ). Although the sample size was small, this suggests that the Xq28 locus is measurably penetrant, and provides independent support for a linkage between male sexual orientation and Xq28 at an overall combined significance level of  $P \approx 0.002–0.02$ . Furthermore, these data clearly exclude the possibility that the observed linkage between Xq28 and male sexual orientation is an artifact due to segregation distortion; if that were true, we would have found more than 50% marker sharing, the exact opposite of the experimental data.

Multipoint interval mapping suggested that the locus related to male sexual orientation lies in the X-specific portion of Xq28, rather than in the telomeric Xq/Yq homology region, and two families with recombination events consistent with this localization were found. By contrast, the previously described families did not contain any unequivocal Xq28 recombinants and were therefore ambiguous with regard to map position within the Xq/Yq–Xq28 region<sup>5</sup>. Simulations indicate that the multipoint regression and maximum likelihood procedures are unbiased in locating a QTL<sup>16,20</sup>. Furthermore, unlike conventional model-dependent multipoint lod score methods, they do not require deliberate misspecification of model parameters to avoid inflating the distance between marker and QTL by confounding recombination with heterogeneity and phenocopies. However, we recognize that the genetic mapping of quantitative traits is complicated by the fact that no single recombination event can be interpreted with certainty, and that a much larger dataset would be required to confirm statistically and refine this location<sup>27–28</sup>.

It is important to recognize that the interpretation of the male linkage results is limited by the highly selective nature of the families studied. Specifically, we analysed only families that were chosen to exhibit a defined level of genetic loading (the presence of two homosexual siblings), the absence of non-sex-linked inheritance (exclusion of families with nonmaternal transmission), and predominant or exclusive homosexuality (exclusion of bisexuals and individuals whose sexual orientation was uncertain). Similarly, heterosexual brothers were analysed only in a limited set of families in which there was a presumption for genetic loading (presence of two Xq28-concordant gay siblings). Therefore, our results cannot be extrapolated to individuals who do not meet our inclusion criteria or to different or more extended family configurations, nor can they be used as

a measure of the overall penetrance of the locus.

Even within the selected population that was studied, the Xq28 region was neither necessary nor sufficient for a homosexual orientation. Within the population at large, its influence is undoubtedly less. Indeed, there was no significant evidence for Xq28 linkage among families that failed to meet the study entrance criteria, and adding these families to the main series decreased the level of homosexual-homosexual gene sharing and increased the level of homosexual-heterosexual gene sharing. Therefore, in order for other investigators to confirm our findings, it will be essential to use the same inclusion and exclusion criteria and to collect detailed and complete family sexual histories and personal information. Although linkage analyses of unselected homosexual populations may be useful for other purposes, such as the search for autosomal loci that influence sexual orientation, they are not suitable for appraising the current findings which were based on the use of a selected study population. Using appropriately selected families improves the accuracy and sensitivity of linkage analysis by decreasing type II error without introducing type I error, and is a well established experimental design in human genetics, especially for sex-linked traits.

**Female findings.** The results for females were notably different than for males (Fig. 3). Analysis of 36 families in which there were two lesbian sisters revealed no significant linkages between Xq28 DNA markers and homosexual orientation. Furthermore, even in those families in which the lesbian siblings were concordant for Xq28 markers, their heterosexual sisters had only a random probability of sharing alleles. Multipoint interval mapping also failed to reveal any correlations between Xq28 markers and the quantitative variables of sexual self-identification, attraction, fantasy and behaviour or the binary variable of sexual orientation. Therefore it appears that the Xq28 locus does not have a major role in individual variations in female sexual orientation. However, we cannot rule out the possibility of a minor contribution because the sample size was limited and because the sib-pair design does not allow testing of father-to-daughter X-linked transmission. Analysis of a larger number and variety of pedigree structures would be useful to resolve this issue and to search for other loci that may correlate with female sexuality.

The simplest explanation of our results is that a locus at Xq28 influences sexual orientation in men but not in women. The fact that most of the sisters of the gay male sib-pairs were heterosexual, regardless of their Xq28 status, supports this interpretation (unpublished results). An alternative model is that the putative Xq28 locus acts in both genders but is recessive in females. Arguing against this idea is the observation that most of the fathers of the lesbian subjects were heterosexual; if a recessive locus were responsible for homosexuality in both sexes, a substantial fraction of the fathers should have been gay. Another possible explanation of the data is that we selected against a sex-shared locus by studying only same-sex sib-pairs. However, preliminary analysis of families with homosexual brother-sister pairs does not support this model (unpublished results).

The finding of a linkage to Xq28 in gay men but not in lesbians suggests that the mechanisms underlying





male and female sexual orientation are at least partially distinct. Family studies have generally shown that gay men have more gay brothers than lesbian sisters and that lesbians have more lesbian sisters than gay brothers, indicating some degree of aetiological independence between male and female sexual orientation<sup>5,6,8,10-13</sup>. Moreover, it is well known that primary sexual development and differentiation involve different gene products in males than in females<sup>3,29</sup>. Our working hypothesis is that male and female sexuality develop along largely independent pathways whereas heterosexuality and homosexuality involve minor variations in gender-specific mechanisms. The development of male homosexuality and heterosexuality appear to involve genetic factors including a locus at Xq28. Whether individual variations in female sexual orientation are influenced by genes remains to be seen.

## Methods

**Subjects.** Families in which there were two non-twin homosexual siblings of the same sex were recruited by advertising in gay and lesbian newspapers, homophile organizations and local clinics. Parents, heterosexual siblings and other family members were encouraged to volunteer. The participants were predominantly white, college educated, and of middle to upper socioeconomic status. The average ages for the siblings were  $34.0 \pm 7.6$  years (mean  $\pm$  standard deviation) in the male study and  $35.3 \pm 7.6$  years in the female study.

Following informed consent, sexual orientation was assessed through a structured interview or questionnaire that included a sexual history and the Kinsey scales of sexual attraction, fantasy, behaviour and self-identification<sup>4-5,14-15</sup>. Subjects were classified as homosexual if they had a predominant or exclusive sexual orientation toward members of the same gender as self-reported during their personal interview or on a questionnaire. Individuals were classified as heterosexual if they had a predominant or exclusive orientation towards members of the opposite gender as determined by interview, questionnaire, or in one case by sibling reports. Individuals who identified as bisexual, defined as roughly equal sexual orientation to both genders, or who gave ambiguous sexual histories or contradictory interview responses, were excluded from the primary analysis. Average Kinsey scores for the included population of subjects were  $5.73 \pm 0.35$  (mean  $\pm$  standard deviation) for homosexual males,  $0.11 \pm 0.25$  for heterosexual males,  $5.41 \pm 0.53$  for homosexual females, and  $0.20 \pm 0.48$  for heterosexual females.

Out of the 40 families interviewed for the male study, 33 were included in the primary linkage analysis. Reasons for the 7 exclusions were as follows: one brother was bisexual or ambiguous rather than homosexual (2 exclusions); there were more than two nonheterosexual brothers (2 exclusions); the father was not exclusively heterosexual (2 exclusions); or there was a homosexual paternal half-brother (1 exclusion). Entrance criteria were set, pedigrees collected, and individual inclusion and exclusion decisions made for each of these families prior to genotype analysis. Families in which the son of a gay sibling is nonheterosexual, or in which there are 2 or more lesbian relatives, would also have been excluded from our analysis of X-linked male sexual orientation, but were not encountered in the present study population. The 33 new families that were included in the main male study consisted of 113 individuals for whom genotypes were determined: 66 gay brothers, 9 heterosexual brothers, 9 sisters, 16 mothers and 13 fathers. In addition, 4 previously described families with Xq28-concordant gay brothers and a heterosexual brother from whom DNA was available were included in determining the penetrance and overall significance of allele sharing ( $\hat{\pi}_{gs}$ ,  $t_{gs}$ , and  $t_{tot}$  in Table 1a), but not in calculating homosexual sib-pair allele sharing ( $\hat{\pi}_{gg}$  and  $t_{gg}$  in Table 1a) or the multipoint analysis (Fig. 2). The 7 families that were excluded from the main study consisted of 24 indi-

viduals for whom genotypes were determined: 13 gay brothers, 4 heterosexual brothers, 2 bisexual brothers, 1 ambiguous brother, 1 sister, 2 mothers and 1 father. The proband for one of these families, which had been excluded because his brother was bisexual, turned out to be heterozygous for two Xq28 markers; this family was not included in Table 2 or Fig. 2.

Out of the 53 families interviewed for the female study, 36 were included in the linkage analysis. Reasons for exclusion were as follows: one sister was bisexual or ambiguous rather than homosexual (9 exclusions); there were more than 2 nonheterosexual sisters (5 exclusions); the lesbian sisters were twins (2 exclusions); or there were 2 gay male relatives (1 exclusion). The 36 families that were included in the female study consisted of 142 individuals for whom genotypes were determined: 72 lesbian sisters, 19 heterosexual sisters, 4 brothers, 15 fathers and 32 mothers.

**Genotyping.** DNA was extracted from peripheral blood by a commercial service (Genetic Design, Greensboro, NC). Samples were genotyped using the following markers and the PCR reaction and analysis conditions described previously<sup>5</sup>: *DXYS154* ( $n$  = number of alleles = 8, het = heterozygosity = 0.81); *DXS1108* ( $n$  = 6, het = 0.75); *F8C* ( $n$  = 2, het = 0.41); *G6PD* ( $n$  = 2, het = 0.20); *DXS52* ( $n$  = 12, het = 0.78); *FRAX-AC2* ( $n$  = 8, het = 0.73); and *DXS548* ( $n$  = 4, het = 0.42).

*JXYQ28*, a newly available marker that was not included in our previous analysis<sup>5</sup>, is based on a 64-bp insertion/deletion polymorphism in the Xq28 X-specific sequences immediately adjacent to the boundary with the Xq/Yq homology region<sup>22</sup>. This polymorphism was detected with the following primers: LLX1: 5'-CAAAATGTGGCATATACAC-3'; LLHOM1: 5'-TAA-TGCTAGATGACACA-3'. 30 cycles of amplification were carried out at 94 °C for 1 min, 92 °C for 1 min, 50 °C for 1 min, and 75 °C for 0.5 min. The products were analysed by 3.5% MetaPhor agarose gel electrophoresis to separate the long allele (239 bp) from the short allele (175 bp). The observed heterozygosity for *JXYQ28* was 0.34. The marker data for the genotyped individuals from both the included and excluded families in the male study are shown in Table 3.

**Data analysis.** Marker allele sharing was analysed by the method of Haseman and Elston<sup>17,21</sup> appropriately modified for X-linkage. The proportion of maternally derived X-linked marker alleles that a sib-pair shares by-descent,  $\hat{\pi}$ , was estimated as follows. If the sibs have different maternal alleles,  $\hat{\pi} = 0$ ; if the sibs have the same maternal allele and the mother is heterozygous,  $\hat{\pi} = 1$ ; if the sibs have the same maternal allele and the heterozygosity of the mother is unknown,  $\hat{\pi} = 1/(1+f)$ , where  $f$  is the frequency of the allele in the population under study; if four sibs have the same maternal allele and the heterozygosity of the mother is unknown,  $\hat{\pi} = (1+3f)/(1+7f)$ ; if the mother is homozygous,  $\hat{\pi}$  is unknown and the pair is not included in the analysis.

Two different multipoint interval mapping approaches were used. The Fulker *et al.* method<sup>20</sup> simultaneously considers markers and their information content to impute the proportion of alleles shared by-descent by a sib-pair ( $\hat{\pi}_q$ ) for any position on the chromosome by a linear regression equation. The Kruglyak and Lander<sup>16</sup> approach computes  $\hat{\pi}_q$  values by a multipoint likelihood approach that uses a new algorithm<sup>30</sup>. In both methods, squared sib-pair trait differences are then regressed on each value of  $\hat{\pi}_q$  along the map to determine the associated t-value by the method of Haseman and Elston<sup>17,21</sup>. Linkage is indicated by negative regression curves and t-scores. Because the two methods gave very similar results, the Fulker *et al.* analyses are shown in Fig. 2 whereas the Kruglyak and Lander analyses are described in the text. Map distances were taken from Wang *et al.*<sup>31</sup>

Significance levels were initially estimated by one-sided t-testing. Because this is an asymptotic test that is most accurate for large sample sizes and normally distributed data, empirical P values for the critical data were determined by simulation



experiments. One thousand to 10,000 replicates were generated by taking the exact family structures, genetic map and loci, fixing the phenotypes at their observed values, and randomly assigning parental genotypes according to the marker allele frequencies. Then parental alleles were chosen at random for transmission to each offspring at a locus at one end of the map and recombination fractions were used to assign the same or different alleles to the other loci. The multipoint maximum likelihood t-score for each replicate was calculated and *P* values were computed from the number of replicates that equalled or exceeded the experimentally observed t-value. Other statistical methods have been described<sup>32</sup>.

**Acknowledgements**

We are grateful to all the participants who made this study possible. We thank A. Wilson for expert technical assistance, M.P. Reeve-Daly and M. Daly for assistance with the statistical analysis, and E. Gershon, C. Klee, E. Lander, M. Lichten, J. Nathans, M. Yarmolinsky and the anonymous referees for critical reviews of the manuscript. This work was supported in part by National Institutes of Health grants HD-11681, HD-27802, and MH-43899 to D.W.F., HG-00017 to L.K., and a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship to S.S.C.

Received 6 March; accepted 20 September 1995.

1. Trip, C.A. *The Homosexual Matrix*. (Signet, New York, 1975).
2. Bell, A.P., Weinberg, M.S. & Hammersmith, S.K. *Sexual Preference: Its Development in Men & Women*. (Indiana University Press, Bloomington, 1981).
3. LeVay, S. *The Sexual Brain*. (MIT Press, Cambridge, MA, 1993).
4. Hamer, D.H. & Copeland, P. *The Science of Desire*. (Simon & Schuster, New York, 1994).
5. Hamer, D.H., Hu, S., Magnuson, V.L., Hu, N. & Pattatucci, A.M.L. A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* **261**, 321–327 (1993).
6. Pattatucci, A.M.L. & Hamer, D. Development and familiarity of sexual orientation in females. *Behav. Genet.* **25**, 407–420 (1995).
7. Rutter, M. Psychiatric genetics: research challenges and pathways forward. *Am. J. med. Genet.* **15**, 185–198 (1994).
8. Bailey, J.M. & Pillard, R.C. A genetic study of male sexual orientation. *Archs. gen. Psychiat.* **48**, 1089–1096 (1991).
9. Bailey, J.M. & Bell, A.P. Familiarity of female & male homosexuality. *Behav. Genet.* **23**, 313–322 (1993).
10. Bailey, J.M. & Benishay, B.A. Familial aggregation of female sexual orientation. *Am. J. Psychiat.* **150**, 272–277 (1993).
11. Bailey, J.M., Pillard, R.C., Neale, M.C. & Agyei, Y. Heritable factors influence sexual orientation in women. *Archs. gen. Psychiat.* **50**, 217–223 (1993).
12. Pillard, R.C. & Weinrich, J.D. Evidence of familial nature of male homosexuality. *Archs. gen. Psychiat.* **43**, 808–812 (1986).
13. Pillard, R.C. The Kinsey scale: is it familial? In *Homosexuality/Heterosexuality: Concepts of Sexual Orientation* (eds McWhirter, D.P., Sanders, S.A. & Reinisch, J.M.) 88–100 (Oxford University Press, Oxford, 1990).
14. Kinsey, A.C., Pomeroy, W.B. & Martin, C.E. *Sexual Behavior in the Human Male*. (W.B. Saunders, Philadelphia, 1948).
15. Kinsey, A.C., Pomeroy, W.B., Martin, C.E. & Gebhard, P. *Sexual Behavior in the Human Female*. (W.B. Saunders, Philadelphia, 1953).
16. Kruglyak, L. & Lander, E.S. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. hum. Genet.* **57**, 439–454 (1995).
17. Haseman, J.K. & Elston, R.C. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3–19 (1972).
18. Cardon, L.R. & Fulker, D.W. The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am. J. hum. Genet.* **55**, 825–833 (1994).
19. Fulker, D.W. & Cardon, L.R. A sib-pair approach to interval mapping of quantitative trait loci. *Am. J. hum. Genet.* **54**, 1092–1103 (1994).
20. Fulker, D.R., Cherny, S.S. & Cardon, L.R. Multipoint interval mapping of quantitative trait loci using sib pairs. *Am. J. hum. Genet.* **56**, 1224–1233 (1995).
21. Tran, L.D., Elston, R.C., Keats, B.J.B. & Wilson, A.F. Sib-pair linkage program. In S.A.G.E. Users Guide, Release 2.2 (Department of Biometry and Genetics, LSU Medical Center, New Orleans, 1994).
22. Freije, D., Helms, C., Watson, M.S. & Donis-Keller, H. Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**, 1784–1787 (1992).
23. Kvaloy, K., Galvagni, F. & Brown, W.R.A. The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Hum. molec. Genet.* **3**, 771–778, 1994.
24. Lander, E.S. & Schork, N.J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994).
25. Pericak-Vance, M.A. et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am. J. hum. Genet.* **48**, 1034–1050 (1991).
26. Cardon, L.R., Smith, S.D., Fulker, D.W., Kimberling, W.J., Pennington, B.F. & DeFries, J.C. Quantitative trait for reading disability on chromosome 6. *Science*, **265**, 276–279 (1994).
27. Boehnke, M. Limits of resolution of genetic linkage studies: implications for the positional cloning of human disease genes. *Am. J. hum. Genet.* **55**, 379–90 (1994).
28. Kruglyak, L. & Lander, E. High-resolution genetic mapping of complex traits. *Am. J. hum. Genet.* **56**, 1212–1223 (1995).
29. Haqq, C.M. et al. Molecular basis of mammalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY. *Science* **266**, 1494–1450 (1994).
30. Kruglyak, L., Daly, M.J. & Lander, E.S. Rapid multipoint analysis of recessive traits in nuclear families, including homozygosity mapping. *Am. J. hum. Genet.* **56**, 519–527 (1995).
31. Wang, L.H., Collins, A., Lawrence, S., Keats, B.J. & Morton, N.E. Integration of gene maps: chromosome X. *Genomics* **22**, 590–604 (1994).
32. Fisher, R.A. *Statistical methods for research workers*, 10th Edition. (Oliver & Boyd, Edinburgh, 1954).